H-002

semi-CNF による手書き日本語文字列認識の弱教師学習
Weakly supervised learning for semi-CNF based handwritten Japanese textline recognition

田中 瑛一† 木村 俊一† 越 裕†
Eiichi TANAKA Shunichi KIMURA Yutaka KOSHI

1. はじめに

本稿では，図 1 に示すような手書き日本語文字列パタンの入力に対して，その読みであるテキスト列を出力する文字列認識を扱う．活字と比較して手書きの文字列はパタンの揺らぎが大きい．また，他の言語と比較して日本語は文字境界が曖昧であり，字種数が膨大である．これらが原因となり手書き日本語文字列の認識は困難な課題となっている．これに対して semi-Markov conditional random fields[1]（semi-CRF）による文字列認識手法[2][3][4]は，日本語や中国語の文字列の性質に適した技術的要素を持つことで優れた成果を示している（2 章）．

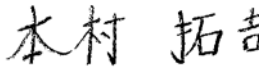

図 1. 氏名文字列の画像サンプル

一般に，semi-CRF 等の機械学習においては，より多くの学習データを利用することで，より良好な認識性能が得られる．この傾向は特に，認識パラメータ数が大きな機械学習において顕著である．このことに対して既存手法[2][3][4]の学習は，テキスト列と文字境界を学習データの教師信号として要求する教師あり学習であるため，学習データ作成のコストが高い，という問題がある．これは特に後者の文字境界の教師信号付与の作業コストが高いためである．また，semi-CRF は認識パラメータ数が小さいため必要となる学習データ量が比較的少ないが，線形であるため認識性能が特徴量設計者である人間の主観に大きく依存する，という問題がある．本稿はこれらの問題を解決するため，以下の 2 つの手法を提案する．

① 文字境界の教師信号不要の学習
② conditional neural fields[5]の適用

①は弱教師学習である．①によって文字境界の教師信号付与の作業コストがなくなり，学習データ作成のコストが下がる．これにより，より多くの学習データが利用可能となる．①の技術的特徴は文字境界を隠れ状態として扱うことである．

†富士ゼロックス株式会社 Fuji Xerox Co., Ltd.

②は semi-CRF による文字列認識への conditional neural fields（CNF）の適用である．すなわち，提案手法の機械学習は semi-CNF と呼べる．充分な学習データ量があるとき，線形である CRF よりも非線形である CNF の方がより良好な認識性能が得られる．②のため，認識パラメータ数が増加し，より多くの学習データが必要となるが，この問題は①によって解決される（3 章）．

手書き日本語文字列画像データによる検証実験の結果，②の CNF 適用による認識性能の改善が示された．また，①の弱教師学習によって同じ学習データ量で教師あり学習と同等の認識性能が得られることが示された（4 章）．

2. 既存手法

2.1 手書き日本語の文字列認識

本稿は semi-CRF による文字列認識[2][3][4]を基盤としている．この手法は日本語や中国語の文字列認識において優れた成果を示しており，これは次の 3 つの技術的要素によって実現されている．

A) 文字の系列による文字列のモデル化
B) 文字境界とテキスト列の semi-CRF による同時推定
C) 外部単文字識別器の利用と異なるクラス間での認識パラメータ共有

まず，A)を説明する．文字列認識では文脈を利用することでより良好な認識性能が得られるため，文字列を単語または文字の系列としてモデル化することが有効である．なお文脈とは，単語や文字の言語・形状・位置に由来する特徴量の前後関係を指す．単語は文字と比較して種別数が非常に多いため，単語を単位とする手法[6]は認識可能な単語の種別数を制限する必要がある．これは特に数値列を対象とする場合に問題となる．また，単語は文字と比較して領域が大きいため手書きによるパタンの揺らぎが大きい．更に，日本語は単語を分かち書きしないため単語境界の特定が困難である．一方，文字は単語と比較して情報量が少なく，その系列が認識のために十分な文脈を持つためには，より多くの系列長が必要となる．しかし，英語等の他の言語と比較すると，日本語の文字の系

列は少ない系列長で比較的豊富な文脈を持つという性質がある．そこで既存手法[2][3][4]は，文字の系列で文字列をモデル化することで，単語由来の困難を回避し，日本語の性質を活用している．

続いて，B)を説明する．文字列内のある文字は，文字列パタンにおける位置を表す文字領域と，そのクラスを表すテキストから成る．A)の文字系列を作成するためには，まず文字境界を特定し，入力の文字列パタンを文字領域に分割する必要がある．しかし，手書き文字列は必ずしも文字の幅や間隔が一様でなく，文字が接触し複数文字がひとつの連結成分から成る場合がある．また日本語文字列は偏や旁のためひとつの文字が複数の連結成分から成る場合がある．更に，文字列に含まれる文字数は一般に未知である．このように，手書き日本語文字列の文字境界は曖昧であるため，単純な手法によって文字領域を特定することは困難である．そこで既存手法[2][3][4]は，まず入力の文字列パタンを文字以下の領域へ過分割し，続いて文字境界（すなわち，文字領域列）とテキスト列を同時に推定する，というアプローチを採用している．テキストの推定結果を文字領域の推定に利用することで，文字境界が曖昧である問題を解決している．これにより，文献[7]のような入力画像サイズの固定と最大文字数の制限を設けることなく，不特定多数文字の認識を実現している．既存手法[2][3][4]は前述の推定に semi-CRF を適用している．semi-CRF は conditional random fields[8]（CRF）の拡張であり，セグメンテーションとラベリングを同時に表す確率モデルである．CRF は識別モデルであるため，hidden Markov model 等の生成モデルと比較してより少ない学習データで良好な認識性能が得られる．また，系列全体の確率を表すため，maximum entropy Markov model[9]等の部分系列の確率を表す識別モデルと比較してより良好な認識性能が得られる．

最後に，C)を説明する．semi-CRF の学習と認識では全文字系列候補に関する計算を行う．この計算は belief propagation（BP）と呼ばれるアルゴリズムによって多項式時間で計算可能である[1]．しかし，文字領域あたり全字種を考慮した場合，日本語の字種数（例えば，氏名認識では約 7,000 種程度が必要）においては処理時間が膨大になる．また，一般的な CRF はクラス（＝字種）の組み合わせ別に認識パラメータを持つ．この組み合わせのため日本語の字種数では認識パラメータ数が膨大なものとなる．例えば，本稿のように隣り合う 2 文字を考慮する場合，組み合わせ数は字種数の 2 乗である．良好な認識性能を得るためには，クラスの組み合わせを網羅し，かつ十分な揺らぎを持つ学習データを準備する必要があるが，これが非現実的なものとなる．そこで既存手法[2][3][4]は，semi-CRF の外部に単文字識別器を持ち，その出力であるテキスト候補を識別スコアで間引くことで処理時間が膨大になる問題を解決している．更に，異なるクラス（の組み合わせ）間で同一の認識パラメータを共有することで認識パラメータ数が膨大になる問題を解決している．なお，このパラメータ共有によって semi-CRF の識別機能が失われるが，これは単文字識別器が補う．

2.2 semi-CRF による文字列認識の学習

一般に，semi-CRF 等の機械学習においては，より多くの学習データを利用することが望ましい．このため，いかに大量の学習データを低コストで作成可能であるかが重要なポイントとなる．

既存手法[2][3][4]はテキスト列と文字境界を教師信号として要求する．このため文字列パタンのデータを学習に利用するためには，人手により教師信号を付与する必要があり，その作業コストが問題となる．また，単純な手法や他の文字列認識器による教師信号の自動付与では，未学習のデータを対象とするため，誤った教師信号が混入する，という問題がある．

これに対して，文献[10][11]ではエントロピーを利用した半教師学習手法が提案されている．半教師学習は教師あり学習と教師なし学習を混合した手法である．教師なし学習の部分が教師信号を要求しないため，学習データ作成のコストが低い．しかし，教師信号を利用しないため教師あり学習と比較して認識性能が劣る傾向がある．また，エントロピー項の係数が制御パラメータとして追加されるため，学習の運用が煩雑になる，という問題がある．

一般に，文字境界は形状的な情報であるためテキスト列と比較して教師信号付与のコストが高い．ゆえに，テキスト列のみが付与されたデータによる学習は，学習データ作成のコスト低下に効果的な方策である．同様のことは画像内の一般物体認識の分野でも検討されている．例えば文献[12] [13] [14] [15] [16]では，画像内の一般物体のクラスが付与されているが位置が付与されていない学習データによる弱教師学習手法が提案されている．しかし，これらは semi-CRF をベースとしておらず，本稿の課題に直接適用することができない．

そこで本稿では，semi-CRF による文字列認識について，テキスト列のみを教師信号として要求し，文字境界の教師信号が不要である弱教師学習手法を提案する．提案手法の弱教師学習により学

習データ作成のコストが下がり，より多くの学習データが利用可能となり，認識パラメータ数増加への対応が容易となる．そこで本稿では，このことを活用し，semi-CRF による文字列認識への CNF 適用を併せて提案する．

3. 提案手法

3.1 semi-CRF による文字列認識の定式化

本稿が想定する文字列認識を定式化する．文字列認識の学習と認識では，まず，図 2 に示すような文字系列候補ラティスを作成する． 文字系列候補ラティスとは，出力しうる全文字系列の候補から成るラティスである．まず，入力の文字列パタンを過分割し文字以下の領域を得る．以下，この領域を，準文字，とする．続いて，連続する準文字の組み合わせから文字領域候補を作成する．続いて，文字領域候補のパタンを単文字識別器に入力しテキスト候補を得る．ある文字領域候補に対して複数のテキスト候補があり，それぞれがラティスのノードである文字候補となる．図 2 上部に準文字を示す．文字領域候補を中括弧で示す．丸角四角形は文字候補を表す．直線は 2 つの文字候補が隣接することを表す． bosとeosは始端と終端を表す模式的なノードである．bosからeosへ隣り合うノードを辿ることで，ある文字系列候補が得られる．

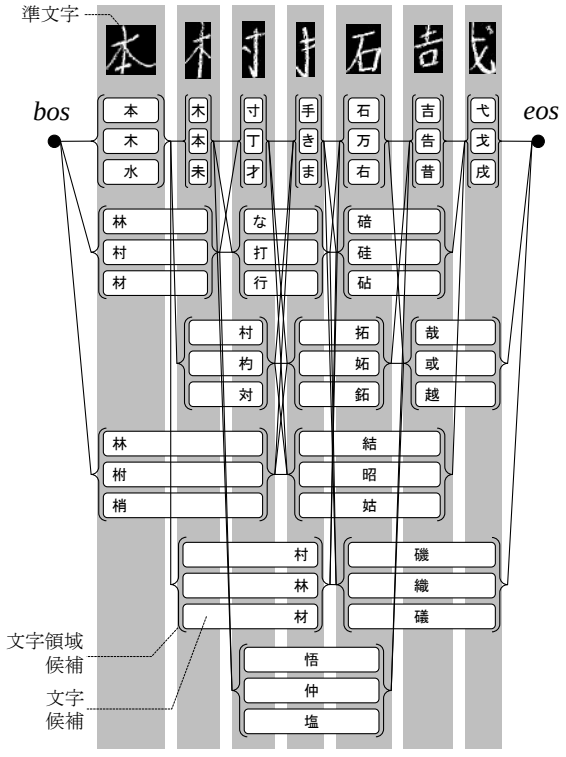

図 2. 準文字列と文字系列候補ラティス

文字系列候補ラティス作成では，文字領域候補を間引く．広すぎる等の冗長な文字領域候補を作成しないことで認識性能と処理速度が改善する．例えば，連続するM個以下の準文字から成る文字領域のみ作成する，という方法がある．図 2 では，$M = 3$としてこの方法を適用している．また同様に，テキスト候補も間引く．例えば，単文字識別の識別スコアの上位K個のテキストを採用する，という方法がある．図 2 では，$K = 3$としてこの方法を適用している．

以下本稿では，準文字列を$X = (x_1, \ldots, x_N)$とする．なお， Nは準文字列のサイズである．また，文字領域列を$S = (s_1, \ldots, s_T)$とする．なお，Tは文字領域列のサイズであり，$T \leq N$であり，未知の文字数である．なお，文字領域は連続する 1 つ以上の準文字から成り，文字列において準文字の重複と過不足がないように作成する．また，テキスト列を$Y = (y_1, \ldots, y_T)$とする．文字列認識の semi-CRF は数式 1 のように書ける．認識は数式 1 の確率が最大となる文字系列(Y^*, S^*)の探索であり， 数式 2 のように書ける．なお，$E(X, Y, S, \Theta)$はエネルギー関数であり，本稿では数式 3 の通り，隣り合う 2 文字候補に関するエネルギーの和で与える．

$$p(Y, S \mid X, \Theta) = \frac{\exp\{-E(X, Y, S, \Theta)\}}{\sum_{(Y', S')} \exp\{-E(X, Y', S', \Theta)\}}$$

数式 1

$$(Y^*, S^*) = \arg\max_{(Y,S)} p(Y, S \mid X, \Theta)$$

数式 2

$$E(X, Y, S, \Theta) = \sum_{t=1}^{t=T} E(y_{t-1}, y_t, s_{t-1}, s_t, X, \Theta)$$

数式 3

3.2 semi-CRF による文字列認識の弱教師学習

提案手法の弱教師学習は文字境界を隠れ状態として扱う．提案手法は，単純な手法や他の文字列認識器による自動付与を行わないため，誤った教師信号の混入がない．また，テキスト列を教師信号として利用するため半教師学習[10][11]と比較してより良好な認識性能が得られる．また，新たな学習の制御パラメータの追加がない．

以下，提案手法の弱教師学習を具体的に説明する．いま，準文字列Xにテキスト列Yのみが教師信号として付与された i.i.d. なデータ$Data = \{(X_d, Y_d)\}_{d=1}^{D}$による学習を考える．学習は$Data$の経験分布と semi-CRF に基づくモデルの確率分布

のカルバック・ライブラー・ダイバージェンスの認識パラメータΘに関する最小化である．これより，数式 4 の損失関数の最小化が導かれる．数式 4 においてlog$p(\Theta)$は事前分布であり学習では正則化として表れる．$loss(X,Y,\Theta)$は$Data$の要素あたりの損失関数であり，数式 5 のように書ける．

$$L(Data,\Theta)=\frac{1}{D}\sum_{(X,Y)\in Data} loss(X,Y,\Theta)-\log p(\Theta)$$

数式 4

$$loss(X,Y,\Theta)=-\log p(Y\mid X,\Theta)$$

数式 5

いま，文字領域列Sが教師信号として与えられないため，$loss(X,Y,\Theta)$を数式 1 の確率で直接与えることができない．そこで，数式 6 に示すように，$p(Y\mid X,\Theta)$を数式 1 の semi-CRF の確率モデルを文字領域列Sで周辺化した確率で与える．

$$p(Y\mid X,\Theta)=\sum_{S'} p(Y,S'\mid X,\Theta)$$

数式 6

学習では，数式 4 の損失関数を最小化するために，その勾配を利用する．すなわち，数式 5 の勾配を利用する．これは数式 7，数式 8 のように導かれる．なおΘ_lはΘの要素を表す．

$$\begin{aligned}&\frac{\partial}{\partial\Theta_l}loss(X,Y,\Theta)\\&=\sum_{S'} p(S'\mid X,Y,\Theta)\frac{\partial}{\partial\Theta_l}E(X,Y,S',\Theta)\\&-\sum_{(Y',S')} p(Y',S'\mid X,\Theta)\frac{\partial}{\partial\Theta_l}E(X,Y',S',\Theta)\end{aligned}$$

数式 7

$$p(S\mid X,Y,\Theta)=\frac{\exp\{-E(X,Y,S,\Theta)\}}{\sum_{S'}\exp\{-E(X,Y,S',\Theta)\}}$$

数式 8

数式 7 が文字領域列Sを要求しないため，本手法は弱教師学習である．なお，以上の定式化は文献[15] [16] [17] [18] [19] [20] [21]等で示される隠れ状態を持つ CRF とほぼ同様である．ただし，提案手法は semi-CRF のセグメンテーションに関する確率変数を隠れ状態として扱う点が異なる．より詳細には，文字列認識における数式 7 の右辺第 1 項と数式 8 の確率の具体的な内容が異なる．

数式 7 の右辺第 2 項は，全文字系列候補に関する期待値計算であり，既存手法と同様である．これは，BP によって高速に計算可能である[1]．一方，数式 7 の右辺第 1 項は，数式 8 の確率に関する期待値計算であり既存手法と異なる．そこで，数式 8 の具体的な内容を考える．数式 8 はテキスト列Yに条件付けられた文字領域列Sの確率を表す．すなわち，テキスト列が固定された複数の文字領域列から成る文字系列候補ラティスに対応する．図 3 にテキスト列が「本村拓哉」に固定された例を示す．以下，このようなラティスを，テキスト列限定ラティス，とする．数式 7 の右辺第 2 項は，図 2 の文字系列候補ラティスを入力とする BP として考えることができる．同様に，数式 7 の右辺第 1 項も，図 3 のテキスト列限定ラティスを入力とする BP として考えることができる．

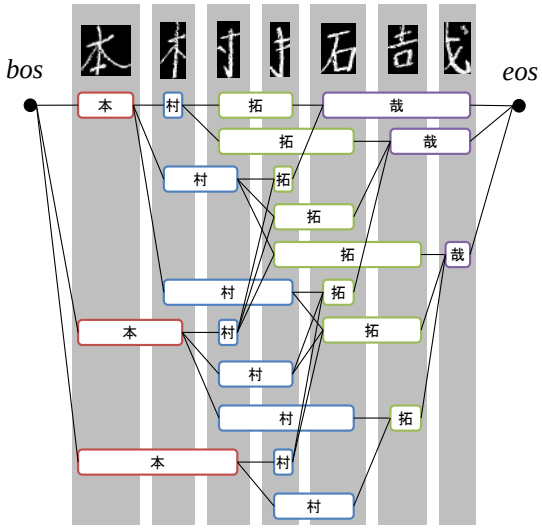

図 3. テキスト列限定ラティス

ただし，数式 7 の右辺第 2 項について，以下のことに注意が必要である．数式 7 の右辺第 2 項は全文字系列候補に関する期待値計算を意味し，右辺第 1 項はその部分集合であるテキスト列が固定された文字系列候補に関する期待値計算を意味する．しかし，テキスト候補の間引きのため，テキスト列限定ラティスは必ずしも文字列候補ラティスの部分グラフにならない．この矛盾を解消するため，提案手法では文字系列候補ラティスとテキスト列限定ラティスの和ラティスを作成する．これを図 4 に示す．図 4 において破線の丸角四角形が新たに追加されたノードを表す．図 4 は図 2 と図 3 の和ラティスであり，テキスト列限定ラティスが部分グラフとなっており，前述の矛盾が解消されている．提案手法では，数式 7 の右辺第 2 項の期待値計算を，和ラティスを入力とする BP で求める．

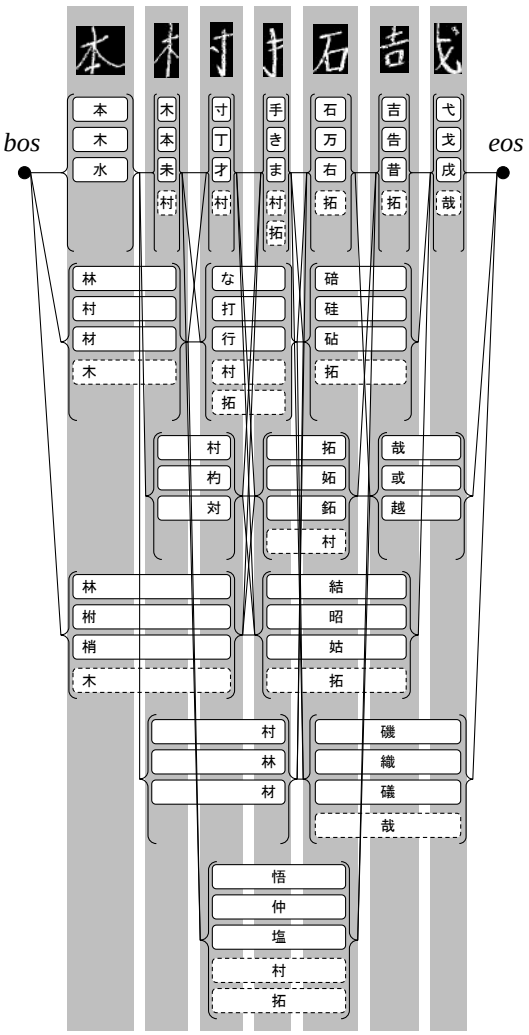

図 4. 和ラティス

3.3 semi-CRF への CNF 適用

提案手法の弱教師学習により，学習データ作成のコストが下がり，より多くの学習データが利用可能となるため，認識パラメータ数増加への対応が容易となる．そこで本稿では認識性能の改善のため，semi-CRF による文字列認識に CNF を適用する．すなわち，semi-CNF による文字列認識を提案する．CRF と異なり，CNF は内部に neural networks（NN）を持つため非線形である．充分な学習データ量があるとき，同じ特徴量であれば，線形である CRF よりも非線形である CNF の方がより良好な認識性能が得られる．これは，CNF の NN の第 1 層に対象に応じて適切な特徴量抽出器が学習され，特徴量設計者である人間の主観への認識性能の依存性が下がるためである．

提案手法における CNF のエネルギー関数を数式 9 に示す．$f_i(y_{t-1}, y_t, s_{t-1}, s_t, X)$は特徴量である．$w_{i,j}$と$\mu_j$は認識パラメータΘの要素である．$\sigma(\cdot)$はシグモイド関数である．$F$は特徴量数である．$J$は NN の隠れ層のサイズである．なお，既存手法[2][3][4]と同様に，全ての認識パラメータについて異なるクラス間で同一の認識パラメータを共有する．また，一般的な CNF では 1 つのノードから成る特徴量と 2 つのノードから成る特徴量を分け，前者のみに NN を適用しているが，提案手法は両者を連結した特徴量ベクトルに対して NN を適用する．これにより，2 つのノードから成る特徴量のみでなく，全ての特徴量を利用した非線形な識別が可能となる．

$$E(y_{t-1}, y_t, s_{t-1}, s_t, X, \Theta) = \sum_{j=1}^{J} \mu_j \sigma\left(\sum_{i=1}^{F} w_{i,j} f_i(y_{t-1}, y_t, s_{t-1}, s_t, X)\right)$$

数式 9

4. 実験結果

4.1 実験データ

実験に使用した手書き文字列画像データの規模を表 1 に示す．本データは筆記者がテキスト列を紙にボールペンで筆記し，これを 2 値スキャンして得られた画像データセットである．データの属性は｛氏名, 住所, 英数記号｝の 3 種類がある．それぞれ学習用と検証用に分かれており，互いに同一の筆記者による文字列がない．なお，氏名と住所の筆記者数は学習用が 279 名，検証用が 31 名である．また，英数記号の筆記者数は学習用が 264 名，検証用が 29 名である．

表 1. 実験データの文字列数

属性	学習用		検証用
	更新用	監視用	
氏名	39,657	4,407	4,915
住所	27,187	3,021	3,339
英数記号	25,306	2,810	2,874

表 2. 学習用データの詳細な情報

属性	字種数	文字列内 平均文字数	文字列内 文字数分散
氏名	2,230	4.0734	0.5671
住所	2,237	23.1904	6.6181
英数記号	84	12.8261	8.1380

また，表 2 に学習データの詳細な情報を示す．属性の“氏名”は日本人の氏名テキスト列から成る．英数記号を含まず，ひらがなカタカナを含む．氏名の画像サンプルを図 1 に示す．属性の“住所”は日本の住所テキスト列から成る．文字列あたりの文字数が多く，英数記号文字を含む．使用される字種数は氏名と同程度である．住所の画像サン

プルを図 5 に示す．属性の“英数記号”は英数記号文字の無秩序な羅列である品番と，メールアドレス，カンマピリオドを含む数字列の計 3 種のテキスト列から成る．字種数は最も少ないが，英語の大文字小文字や「1（数字）」と「l（英字）」と「)」といった手書きによってほぼ識別不可能となる文字が多く筆記される．以下，このような文字のセットを，類似文字，とする．英数記号の画像サンプルを図 6 に示す．

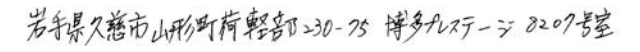

図 5. 住所文字列の画像サンプル

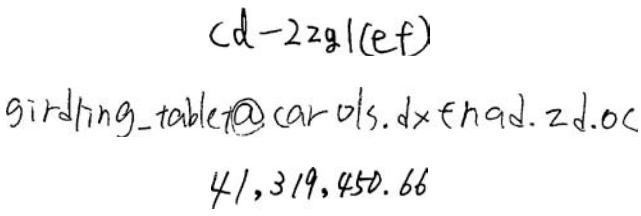

図 6. 英数記号文字列の画像サンプル

4.2 その他実験条件

実験において，過分割手法は最短経路の収束による準文字切り出し手法[22]を利用した．また，文字領域の作成規則は，最大準文字数が$M = 6$ であることと，文字幅が文字列の高さの 2.5 倍以下であることとした．これは，ひとつの文字領域として充分に大きい値である．また単文字識別器は，視覚の方位交差抑制性を持つ convolutional neural networks による手法[23]（CNN）を利用し，識別スコアについて上位 10 位（すなわち，$K = 10$）までをテキスト候補とした．

semi-CRF/CNF の特徴量は，文字の幅，高さ，重心位置，文字間の距離から成る特徴量と，CNN の識別スコアと，テキスト列の出現確率を採用した．なお，テキスト列の出現確率は uni-gram と bi-gram の 2 つである．uni-gram とは，テキストy_tの確率$p(y_t)$を表す．また，bi-gram とは，テキストy_tの条件付き確率$p(y_t \mid y_{t-1})$を表す．以下ではこれらを合わせて，n-gram，とする．なお，本実験の CNN は「い」と「ぃ」や，「C」と「c」など，大きさが異なるが形状が同一の文字をひとつのクラスとしている．以下，このような文字のセットを，同形異字，とする．ゆえに，本実験の CNN には同形異字である幼音促音や一部の大文字小文字を識別する機能がない．文字列候補ラティス作成においては，CNN がひとつのクラスとした同形異字を複数のテキスト候補に展開しノードを作成する．これらのノードの特徴量は n-gram 以外が全て同じ値となる．

また，CNN と n-gram は，表 1，表 2 以外のデータで学習済みの，各属性に特化したものを使用した．氏名と住所の CNN は，外字の存在や，実践では氏名欄や住所欄に法人名が筆記される場合もあることなどを想定し，十分大きな字種数（= 7,317 種）に対応したものを使用した．一方，英数記号では英数記号文字に限定した字種数（= 84 種）に対応したものを使用した．また，英数記号では基本的に無秩序に英数記号文字が並ぶため，n-gram は不使用（常に 0）とした．

semi-CRF/CNF の学習は，表 1 の更新用データを利用して，stochastic gradient descent（SGD）と慣性項によって行った．SGD のバッチサイズは 64 である．認識パラメータの初期値は[−1,1]の一様分布の乱数で与え，学習率と慣性率は常に一定とした．なお，正則化はない．学習の収束や過学習を監視するため，表 1 の監視用データを用いて，認識パラメータを 200 回更新する毎に数式 4 の損失関数を計算した．

以上の設定は，全ての手法において共通とした．ただし，CNF の認識パラメータ数は J=16 とした．

4.3 属性別 認識性能比較 実験

文字列認識の認識性能の評価結果を表 3，表 4，表 5 に示す．評価指標について，Text はテキスト列の認識性能を，Segmentation は文字領域列の認識性能を表す．それぞれの系列について，数式 2 で得られる出力系列と正解系列の編集距離を計算し，precision（= 一致系列長÷出力系列長）と recall（= 一致系列長÷正解系列長）を算出した．なお系列長は，検証用データの全ての文字列に関する総和である．また，手法のモデルについて，「CRF」は semi-CRF を，「CNF」は semi-CNF を表す．また，手法の学習について，「FULL」は教師あり学習を，「WEAK」は提案手法の弱教師学習を表す．すなわち，本実験において FULL と WEAK の違いは，後者が文字境界を教師信号として利用しないことのみである．

認識性能を評価した認識パラメータは，SGD について，氏名は 50 万回（およそ，807 epoch），住所は 50 万回（およそ 1177 epoch），英数記号は 100 万回（およそ，2529 epoch）の認識パラメータ更新を行って得られたものである．なお，1 epoch は全ての更新用データを一巡したことを表す．なお，全ての手法と属性において学習の収束が確認され，過学習は確認されなかった．

全ての属性において，CNF 適用により認識性能が改善している．氏名と住所が英数記号よりも改善幅が大きく，氏名と住所では英数記号よりも高

い非線形性が必要であったといえる．

英数記号は他の 2 つと比較して Text の認識性能が低い．これは，同形異字と類似異字が原因である．これらは形状による識別が困難であるため，言語的な特徴量である n-gram によって識別する必要がある．しかし本実験の英数記号では，n-gram が不使用のため，同形異字と類似異字の識別は原理的に困難であった．すなわち，英数記号に対しては，本実験の特徴量が不充分なものであったといえる．

表 3. 認識性能比較，氏名

手法		Text		Segmentation	
モデル	学習	precision	Recall	precision	recall
CRF	FULL	0.9171	0.9199	0.9518	0.9547
CNF	FULL	0.9828	0.9818	0.9903	0.9893
CNF	**WEAK**	0.9804	0.9810	0.9891	0.9897

表 4. 認識性能比較，住所

手法		Text		Segmentation	
モデル	学習	precision	Recall	precision	recall
CRF	FULL	0.8709	0.8641	0.9336	0.9262
CNF	FULL	0.9726	0.9704	0.9847	0.9825
CNF	**WEAK**	0.9747	0.9728	0.9856	0.9837

表 5. 認識性能比較，英数記号

手法		Text		Segmentation	
モデル	学習	precision	Recall	precision	recall
CRF	FULL	0.7993	0.7955	0.9827	0.9780
CNF	FULL	0.8148	0.8121	0.9912	0.9878
CNF	**WEAK**	0.8144	0.8109	0.9923	0.9880

教師あり学習と弱教師学習の認識性能の比較について，氏名と英数記号では教師あり学習の方が僅かに高い．一方，住所は弱教師学習の方が僅かに高い．しかし，CRF の認識性能との差と，弱教師学習によってより多くの学習データが利用可能となることを加味すれば，この差は無視できるほど小さいといえる．すなわち，提案手法の弱教師学習によって，同じ学習データ量で教師あり学習と同等の認識性能が得られることが示された．

ただし，監視用データに対する損失関数の最小化度合いは，教師あり学習よりも弱教師学習が劣る傾向があることが確認された．氏名データにおいて同様の学習を 8 回試行した結果を図 7 に示す．図 7 の縦軸は数式 4 の損失関数であり，横軸はパラメータの更新回数である．左右のグラフで縦軸のレンジを合わせている．色の違いは試行の違いを表す．損失関数の計算ではどちらも文字領域列を利用しているため，図 7 は学習手法の違いによる損失関数の最小化度合いの違いとして見ることができる．この結果から，提案手法の弱教師学習は，損失関数の最小化度合いは教師あり学習に劣るが，それは文字単位の認識性能として顕在化しない程度であるといえる．

4.4 データ量別 認識性能比較 実験

提案手法の弱教師学習の効用は，より多くの学習データが利用可能となることである．そこで，氏名データを利用して学習データ量と認識性能の関係を評価した．

表 6，表 7，表 8 にデータ量に対する認識性能の評価結果を示す．DataVolume は表 1 の氏名の学習データのうち，学習に利用した量の割合を表す．また，相関係数はデータ量と認識性能の相関係数である（表 3 の結果も 100%として加味している）．表 6 に示す通り，semi-CRF はデータ量と認識性能の相関が小さく，学習データ量の増加による改善がない状態といえる．一方，表 7，表 8 に示す通り，semi-CNF はデータ量と認識性能の相関が大きい．これは，認識パラメータの大きい semi-CNF では，その潜在的な認識性能を引き出すため，より多くの学習データ量が必要となっていることを示している．

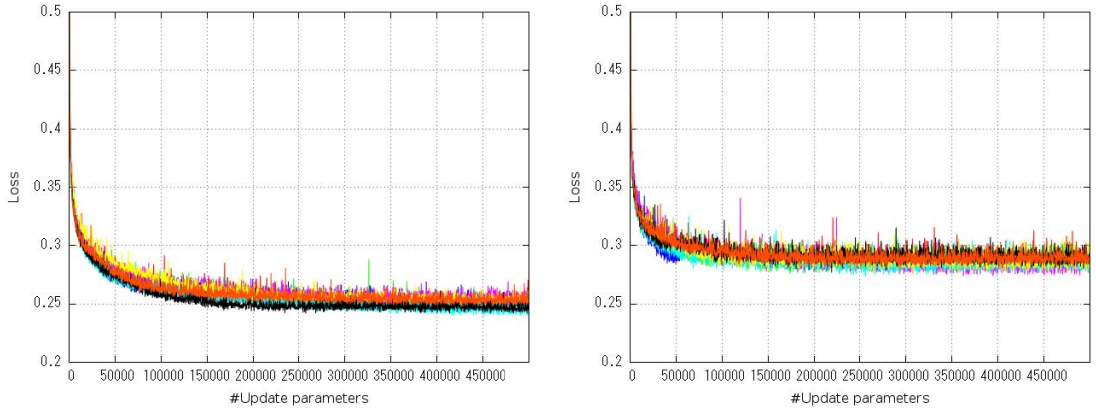

図 7. 学習の監視曲線，CNF，氏名データ，（左： FULL，右： WEAK），縦軸は数式 4 の損失関数，横軸は SGD による認識パラメータの更新回数．色の違いは試行の違いを表す．

表 6. データ量と認識性能，semi-CRF，FULL

Data Volume	Text		Segmentation	
	Precision	recall	precision	recall
10%	0.9162	0.9189	0.9520	0.9547
20%	0.9213	0.9168	0.9567	0.9520
30%	0.9190	0.9221	0.9533	0.9564
60%	0.9221	0.9206	0.9559	0.9543
80%	0.9204	0.9205	0.9546	0.9547
相関係数	0.048926	0.363739	-0.154710	0.175264

表 7. データ量と認識性能，**semi-CNF**，FULL

Data Volume	Text		Segmentation	
	Precision	recall	precision	recall
10%	0.9767	0.9758	0.9878	0.9869
20%	0.9793	0.9793	0.9890	0.9889
30%	0.9808	0.9796	0.9896	0.9884
60%	0.9798	0.9804	0.9886	0.9892
80%	0.9820	0.9820	0.9902	0.9902
相関係数	0.855931	0.867783	0.767818	0.758656

表 8. データ量と認識性能，**semi-CNF**，**WEAK**

Data Volume	Text		Segmentation	
	Precision	recall	precision	recall
10%	0.9752	0.9756	0.9869	0.9873
20%	0.9783	0.9769	0.9877	0.9864
30%	0.9788	0.9793	0.9877	0.9882
60%	0.9820	0.9814	0.9897	0.9891
80%	0.9827	0.9824	0.9905	0.9902
相関係数	0.789980	0.871168	0.819956	0.897278

5. まとめ

一般に，機械学習では，より多くの学習データを利用することで，より良好な認識性能が得られる．この傾向は，本稿の CNF 適用の例のように，特に認識パラメータ数の大きな機械学習において顕著である．これに対して本稿では，文字境界不要の文字列認識の弱教師学習手法を提案した．文字境界の教師信号付与のコストがなくなり，学習データ作成のコストが下がるため，提案手法によってより多くの学習データが利用可能となる．

氏名の手書き日本語文字列画像データを用いた実験により文字単位の recall を評価した結果，教師あり学習において semi-CRF が 91.99% であったのに対して semi-CNF は 98.18% を示し，CNF 適用による認識性能の改善が示された．また，semi-CNF において弱教師学習は 98.10% を示し，提案手法の弱教師学習によって同じ学習データ量の教師あり学習と同等の認識性能が得られることが示された．

参考文献

[1] S. Sarawagi and W. W. Cohen, “Semi-Markov conditional random fields for information extraction,” *In Proc. NIPS 2004*, pp 1185-1192, (2004).

[2] X. D. Zhou, C. L. Liu, and M. Nakagawa, “Online handwritten Japanese character string recognition using conditional random fields,” *In Proc. ICDAR 2009*, Washington, DC, USA, pp. 521‑525, (2009).

[3] X. D. Zhou, Y. M. Zhang, F. Tian, H. A. Wang, and C. L Liu, “Minimum-risk training for semi-Markov conditional random fields with application to handwritten Chinese/Japanese text recognition,” *Pattern Recognition*, Vol. 47, NO. 5, pp. 1904-1916, (2014).

[4] X. D. Zhou, D. H. Wang, F. Tian, C. L. Liu, and M. Nakagawa, “Handwritten Chinese/Japanese text recognition using semi-Markov conditional random fields,” *IEEE Trans, Pattern Analysis and Machine Intelligence*, Vol. 35, No. 10, pp. 2413-2426, (2013).

[5] J. Peng, L. Bo, and J. Xu, “Conditional neural fields,” *In Proc. NIPS 2009*, pp. 1419-1427, (2009).

[6] V. Goel, A. Mishra, K. Alahari and C. V. Jawahar, “Whole is greater than sum of parts: Recognizing scene text words.” *In Proc. ICDAR 2013*, Washington, DC, USA, (2013).

[7] M. Jaderberg, K. Simonyan, A. Vedaldi and A. Zisserman, “Deep Structured Output Learning For Unconstrained Text Recognition,” *In Proc. ICLR 2015*, May 7-9, San Diego, CA, USA, (2015).

[8] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” *In Proc. ICML 2001*, San Francisco, CA, USA, pp. 282‑289, (2001).

[9] A. McCallum, D. Freitag and F. Pereira, “Maximum entropy Markov models for information extraction and segmentation,” *In Proc. ICML 2000*, Stanford, California, pp. 591–598, (2000).

[10] Y. Grandvalet and Y. Bengio, “Semi-supervised learning by entropy minimization,” *In Proc. NIPS 2004*, No.17, pp.529-536, (2004).

[11] F. Jiao, S. Wang, C.H. Lee, R. Greiner and D. Schuurmans, “Semi-supervised conditional random fields for improved sequence segmentation and labeling,” *In Proc. COLING･ACL 2006*, Sydney, Australia, July 17–21, pp.209-216, (2006).

[12] C. Galleguillos, B. Babenko, A. Rabinovich and S. Belongie, “Weakly supervised object localization with stable segmentations,” *In Proc. ECCV 2008*, pp.193-207, (2008).

[13] A. Vezhnevets and J. M. Buhmann, “Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning,” *In Proc. CVPR 2010*, pp.3249-3256, (2010).

[14] M. Oquab, L. Bottou, I. Laptev and J. Sivic, “Weakly supervised object recognition with convolutional neural networks,” (2014).

[15] Z. J. Zha, X. S. Hua, T. Mei, J. Wang, G. J. Qi, and Z. Wang, “Joint multi-label multi-instance learning for image classification,” *In Proc. CVPR 2008*, pp.1-8, (2008).

[16] D. Duvenaud, B. Marlin and K. Murphy, “Multiscale conditional random fields for semi-supervised labeling and classification,” *In Proc. CRV 2011*, pp.371-378, (2011).

[17] A. Quattoni, M. Collins and T. Darrell, “Conditional random fields for object recognition,” *In Proc. NIPS 2004*, pp.1097-1104, (2004).

[18] A. Quattoni, S. Wang, L. P. Morency, M. Collins and T. Darrell, “Hidden conditional random fields,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.29, No.10, pp.1848-1852, (2007).

[19] L. P. Morency, A. Quattoni and T. Darrell, “Latent-dynamic discriminative models for continuous gesture recognition,” *In Proc. CVPR 2007*, pp.1-8, (2007).

[20] M. Mahajan, A. Gunawardana and A. Acero, “Training algorithms for hidden conditional random fields, “*In Proc. ICASSP 2006*, (2006).

[21] Y. H. Sung and D. Jurafsky, “Hidden conditional random fields for phone recognition,” *In Proc. ASRU 2009*, pp.107-112, (2009).

[22] 田中瑛一, “最短経路の収束を利用した文字切り出し方式の提案,” 第 14 回 画像の理解・認識シンポジウム, *MIRU2011*, Sep, (2011).

[23] 関野雅則, 木村俊一, 越裕, “視覚情報処理モデルに基づいて改良した畳込みニューラルネットワーク文字認識,”人工知能学会, *JSAI2013*, (2013).